

**PKM PENINGKATAN KOMPETENSI DOSEN MELALUI PELATIHAN DATA
MINING DENGAN TOOLS RAPID MINER PADA PROGRAM PASCASARJANA
PROGRAM STUDI S2 ILMU KESEHATAN MASYARAKAT
UNIVERSITAS HANG TUAH PEKANBARU**

Eka Sabna¹, Azlina²

Universitas Hang Tuah Pekanbaru^{1,2}

e-mail korespondensi : es3jelita@yahoo.com¹

ABSTRACT

Technological developments in the Industry 4.0 era open up enormous opportunities in data collection and processing. Currently, health data makes up around 30% of all global data, and by 2025, this figure will reach 36%. The ability to understand such segmented data can provide a major strategic advantage to medical organizations everywhere. This large amount of data can be analysed in research using various approaches, including the Data Mining Approach. Data Mining can be applied to find knowledge patterns from patient profiles and health history data (patient history data). The knowledge gained can be used for analysis and decision-making, including predicting the type of disease, determining the pattern of disease spread, and assessing the effectiveness of treatment. Some Lecturers in the Master of Public Health study program are unfamiliar with the basic concepts of data analysis, particularly those related to data mining. This activity aims to provide lecturers in the Master of Public Health Study Program with knowledge about analysing health data using a data mining approach. The Data Mining technique discussed is the prediction of diabetes using the Decision Tree Algorithm. The data used was obtained from public data, namely Kaggle, with 786 data points consisting of 9 attributes. This data set comes from the National Institute of Diabetes and Digestive and Kidney Diseases. The goal of this data set is to diagnostically predict whether a patient has diabetes, based on certain diagnostic measurements included in the data set.

Keywords: *Diabetes, Data Mining, Prediction, Lecturers, Decision Tree*

ABSTRAK

Perkembangan teknologi di era Industri 4.0 membuka peluang yang sangat besar dalam pengumpulan dan pengolahan data. Saat ini, data kesehatan memproduksi sekitar 30% dari seluruh data global, dan pada tahun 2025, angka ini akan mencapai 36%. Kemampuan untuk memahami data yang tersegmentasi tersebut dapat memberikan keuntungan strategis yang besar bagi organisasi medis di mana pun. Jumlah data yang besar ini dapat dilakukan penelitian-penelitian dengan berbagai pendekatan diantaranya adalah dengan menggunakan Pendekatan *Data Mining*. *Data Mining* dapat diterapkan untuk menemukan pola pengetahuan dari profil pasien dan data riwayat kesehatannya (History data pasien). Pengetahuan yang diperoleh dapat digunakan analisis dan pengambilan keputusan, antara lain untuk memprediksi jenis penyakit, mengetahui pola penyebaran penyakit, dan melihat efektifitas pengobatan. Sebagian Dosen Program Studi S2 Kesehatan Masyarakat belum banyak mengetahui konsep dasar Analisis Data dengan menggunakan konsep *Data Mining*. Kegiatan ini bertujuan untuk memberi pengetahuan tentang analisis data-data kesehatan dengan menggunakan Pendekatan *Data Mining* kepada Dosen-dosen Program Studi S2 Kesehatan Masyarakat. Teknik *Data Mining* yang dibahas adalah Prediksi penyakit diabetes menggunakan Algoritma Decision Tree. Data yang digunakan diperoleh dari data publik yaitu Kaggle dengan 786 data yang terdiri dari 9 atribut. Kumpulan data ini berasal dari Institut Nasional Diabetes dan Penyakit Pencernaan dan Ginjal. Tujuan dari kumpulan data ini adalah untuk memprediksi secara diagnostik apakah seorang pasien menderita diabetes, berdasarkan pengukuran diagnostik tertentu yang termasuk dalam kumpulan data tersebut.

Kata Kunci : Diabetes, Data Mining, Prediksi, Dosen, Pohon Keputusan

PROSIDING SEMINAR NASIONAL

Hasil Penelitian dan Pengabdian Masyarakat

PENDAHULUAN

Saat ini, data kesehatan memproduksi sekitar 30% dari seluruh data global, dan pada tahun 2025, angka ini akan mencapai 36%. Kemampuan untuk memahami data yang tersegmentasi tersebut dapat memberikan keuntungan strategis yang besar bagi organisasi medis di mana pun. Perkembangan teknologi di era Industri 4.0 membuka peluang yang sangat besar dalam pengumpulan dan pengolahan data. Jumlah data yang besar ini dapat dilakukan penelitian-penelitian dengan berbagai pendekatan diantaranya adalah dengan menggunakan Pendekatan *Data Mining* yang merupakan salah satu perkembangan teknologi di bidang komputer (Husen et al., 2022) .

Teknik *Data Mining* menjadi semakin populer dan semakin penting, terutama di bidang medis. *Data Mining* dapat diartikan sebagai proses pencarian pola dan tren yang sebelumnya tidak terlihat pada data dalam jumlah yang sangat besar. *Data Mining* dapat diterapkan dalam melakukan analisis data-data kesehatan. *Data Mining* adalah metode pengolahan data untuk mencari pola tersembunyi dalam data tersebut sehingga pola tersebut dapat dipakai sebagai pengetahuan (Binus, 2021) (Jiawei, n.d.). Secara umum, penerapan *Data Mining* dalam bidang kesehatan diantaranya adalah untuk Peningkatan pengambilan keputusan klinis, Peningkatan akurasi diagnosis, Peningkatan efisiensi pengobatan, Menghindari interaksi obat dan makanan yang berbahaya , Hubungan pelanggan yang lebih baik (dqlab, 2022) .

Tools yang digunakan untuk pelatihan ini adalah Rapidminer. RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap *Data Mining*, *text mining* dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik (Setianingsih, 2015).

Program Studi S2 Kesehatan Masyarakat Fakultas Kesehatan Universitas Hang Tuah Pekanbaru merupakan salah satu Program Studi di Universitas Hang Tuah Pekanbaru. Dosen Prodi S2 Kesehatan Masyarakat selama ini dalam melakukan proses analisis data menggunakan Statistik dan perangkat lunak SPSS. Dosen-dosen belum banyak mengetahui konsep dasar Analisis Data dengan menggunakan konsep *Data Mining*.

Pelatihan ini memberikan alternative lain dalam analisis data yaitu dengan pendekatan *Data Mining* dan tools Rapidminer yang saat ini menjadi pilihan peneliti-peneliti bidang kesehatan untuk melakukan analisis data seperti Prediksi, Klasifikasi, Pengelompokkan (*Clustering*) dan Asosiasi Data (Ucha Putri et al., 2021).

Program Studi Kesehatan Masyarakat merupakan salah satu Program Studi yang ada Universitas Hang Tuah Pekanbaru. Mahasiswa Program Studi Kesehatan Masyarakat belum banyak mengetahui konsep dasar Analisis Data dengan menggunakan konsep *Data Science* ini oleh karena itu maka di usulkan kegiatan pelatihan Analisa Data-data Kesehatan dengan menerapkan *Data Science*. Kegiatan ini bertujuan untuk memberi pengetahuan tentang analisis data-data kesehatan dengan menggunakan *Data Science* kepada Mahasiswa Program Studi Kesehatan

PROSIDING SEMINAR NASIONAL

Hasil Penelitian dan Pengabdian Masyarakat

Masyarakat.

Tujuan Kegiatan ini adalah memberikan Pelatihan tentang *Data Science* agar mahasiswa dapat menganalisa data-data kesehatan dapat menggunakan pendekatan *Data Science*. Manfaat kegiatan ini adalah memberikan dan menambah pengetahuan Mahasiswa dalam menganalisa data-data kesehatan sehingga mahasiswa program studi kesehatan masyarakat dapat menggunakan *Data Science* sebagai alternative lain dimana selama ini proses analisis data hanya menggunakan analisis statistik.

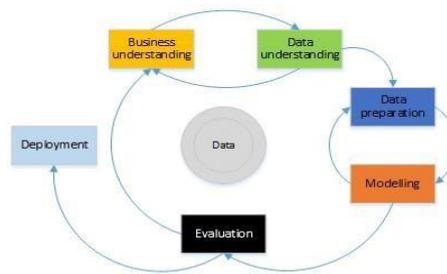
METODE

Berdasarkan hasil identifikasi masalah, tim pengusul melakukan tahapan-tahapan untuk menyelesaikan permasalahan pada mitra. Metode pelaksanaan yang dalam kegiatan PKM ini dibagi beberapa tahapan . Adapun tahapan yang dilakukan dalam pelaksanaan PKM ini adalah sebagai berikut:



Gambar 1. Metode Pelaksanaan Kegiatan PKM

1. Perizinan
Proses pengurusan perizinan dilakukan bersama dengan pihak Program Studi Kesehatan Masyarakat.
2. Penetapan Waktu dan Tempat
Koordinasi Tim PKM dengan Program Studi . Tim melakukan survey, observasi dan melakukan diskusi serta mengatur agenda kegiatan, berkoordinasi terkait perihal waktu dan tempat.
3. Persiapan.
Persiapan Materi dan Peserta kegiatan PKM
 - 1) Menyiapkan topik atau materi yang akan disampaikan dalam kegiatan PKM. Penetapan materi dan peserta pelatihan dilakukan oleh tim PKM untuk mencapai tujuan dari PKM.
 - 2) Peserta
Peserta Mahasiswa Program Studi Kesehatan Masyarakat
4. Pelaksanaan.
Pelatihan kepada Mahasiswa , materi yang akan disampaikan mengikut metode /langkah langkah dalam Data Mining yaitu CRIPS-DM. CRISP-DM memiliki 6 fase tahapan yaitu :(Hotz, 2024; Kase Sukesha, 2022)



Gambar 2. Metode Pelaksanaan Kegiatan PKM

- 1) *Business Understanding* (Pemahaman Bisnis)
 - a. Menganalisa Kebutuhan
 - b. Menentukan Tujuan dan Rencana
 - c. Rumusan masalah data mining.
- 2) *Data Understanding* (Pemahaman Data)
Pengumpulan data, mengenali lebih lanjut data yang akan digunakan .
- 3) *Data Preparation* (Persiapan/Pengolahan Data)
Merupakan tahapan yang sangat penting untuk merancang model Prediksi. Untuk meningkatkan kualitas data yang akan dianalisis, perlu dilakukan langkah-langkah *Preparation Data* yaitu (3) :
 - a. Pembersihan Data (*Data Cleaning*)
Data yang baru saja dikumpulkan mungkin memiliki banyak bagian yang tidak sesuai dan beberapa bagian yang hilang sehingga diperlukan proses pembersihan data.
 - b. Transformasi Data (*Data transformation*)
Data transformation digunakan untuk mengubah data dalam bentuk yang sesuai.
 - c. Mereduksi Data (*Data Reduction*)
Bertujuan untuk meningkatkan efisiensi penyimpanan serta mengurangi biaya penyimpanan dan analisis data.
- 4) *Modelling* (Pemodelan)
Menggunakan model yang dihasilkan. Keakurasian suatu prediksi ditentukan oleh seberapa besar penyimpangan atau kesalahan ini, yang terjadi antara data yang diprediksi dengan data yang sebenarnya atau data actual (4).
- 5) *Evaluation* (Evaluasi)
Evaluasi yaitu melakukan validasi merupakan tahapan yang sangat penting pemodelan, untuk melihat sejauh mana kehandalan model yang akan digunakan dalam hal pengambilan keputusan (5).
- 6) *Deployment* (Penyebaran)
Menggunakan model yang dihasilkan dan pembuatan laporan .

PROSIDING SEMINAR NASIONAL

Hasil Penelitian dan Pengabdian Masyarakat

5. Evaluasi Kegiatan dan Pelaporan

a. Evaluasi

Setelah kegiatan PKM dilaksanakan maka perlu dilakukan evaluasi dengan cara :

- 1) Mengukur keefektivan dan efisiensi dari pelatihan yang dilakukan.
- 2) Keberlanjutan Kerjasama dengan mitra akan terus terjalin setelah kegiatan PKM, seperti ikut serta dalam melakukan pelatihan dan pendampingan terkait proses analisis data –data kesehatan

b. Laporan

Tahap terakhir dalam kegiatan ini yaitu menyusun pelaporan mulai dari awal kegiatan sampai tahap evaluasi. Laporan ini dapat digunakan sebagai referensi.

HASIL

Pelatihan yang akan dilaksanakan dimana Peserta akan diberikan pengetahuan mengenai Data Mining. Materi pelatihan kepada Dosen akan disampaikan mengikuti metode/langkah-langkah dalam Data Mining yaitu CRIPS-DM. Pada Gambar 2 memperlihatkan tahapan CRIPS-DM :

a. Bussiness Understanding (Pemahaman Bisnis)

Kumpulan data ini berasal dari National Institute of Diabetes and Digestive and Kidney Diseases.

Tujuan :

Menemukan pola dari kumpulan data pasien yang menderita Diabetes untuk mem- prediksi penyakit Diabetes.

b. Data Understanding (Pemahaman Data)

Informasi tentang atribut pada data set :

- Pregnancies (Kehamilan) : To express the Number of pregnancies (Untuk menyatakan jumlah kehamilan)
- Glucose (Glukosa) : To express the Glucose level in blood (untuk menyatakan kadar Glukosa dalam darah)
- BloodPressure (Tekanan Darah) : To express the Blood pressure measurement (untuk menyatakan pengukuran tekanan darah).
- SkinThickness: To express the thickness of the skin (untuk menyatakan ketebalan kulit).
- Insulin: To express the Insulin level in blood (untuk menyatakan kadar insulin dalam darah)
- BMI: To express the Body mass index (untuk menyatakan indeks massa tubuh)
- DiabetesPedigreeFunction: To express the Diabetes percentage (untuk menyatakan persentase Diabetes)
- Age: To express the age (untuk mengekspresikan umur)
- Outcome: To express the final result 1 is Yes and 0 is No (untuk menyatakan hasil akhir)

PROSIDING SEMINAR NASIONAL

Hasil Penelitian dan Pengabdian Masyarakat

c. Data Preparation (Persiapan Data)

Data set: <https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset>

Row No.	Outcome	Pregnancies	Glucose	BloodPress...	SkinThicke...	Insulin	BMI	Diabetes
1	1	6	148	72	35	0	33.600	0.627
2	0	1	85	66	29	0	26.600	0.351
3	1	8	183	64	0	0	23.300	0.672
4	0	1	89	66	23	94	28.100	0.167
5	1	0	137	40	35	168	43.100	2288
6	0	5	116	74	0	0	25.600	0.201
7	1	3	78	50	32	88	31	0.248
8	0	10	115	0	0	0	35.300	0.134
9	1	2	197	70	45	543	30.500	0.158
10	1	8	125	96	0	0	0	0.232
11	0	4	110	92	0	0	37.600	0.191
12	1	10	168	74	0	0	38	0.537
13	0	10	136	80	0	0	37.100	1.441

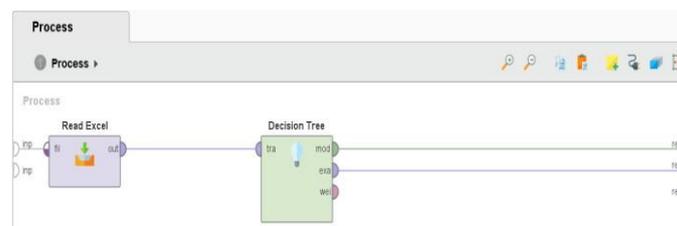
Gambar 3. Preprocessing Data

Name	Type	Missing	Statistics	Filter (9 / 9 attributes)	Value
Label Outcome	Binominal	0	Negative 1 Positive 0		Value 0 (5)
Pregnancies	Real	0	Min 0 Max 17		Average 3.84
Glucose	Real	0	Min 0 Max 199		Average 120
BloodPressure	Real	0	Min 0 Max 122		Average 69.1
SkinThickness	Real	0	Min 0 Max 99		Average 20.5
Insulin	Real	0	Min 0 Max 846		Average 79.7
BMI	Real	0	Min 0 Max 67.100		Average 31.5

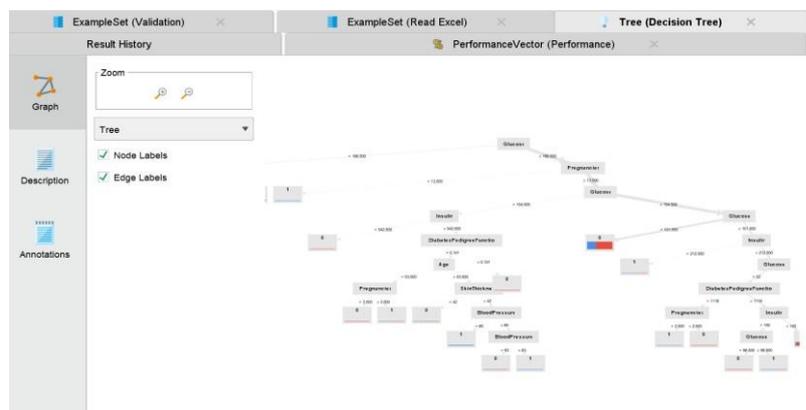
Gambar 4. Proses Cleaning Data

d. Modelling (Pemodelan)

Pemodelan yang digunakan adalah Decisosn Tree. Gambar 5 pemodelan Decision Tree dengan Rapidminer.



Gambar 5. Proses Pemodelan



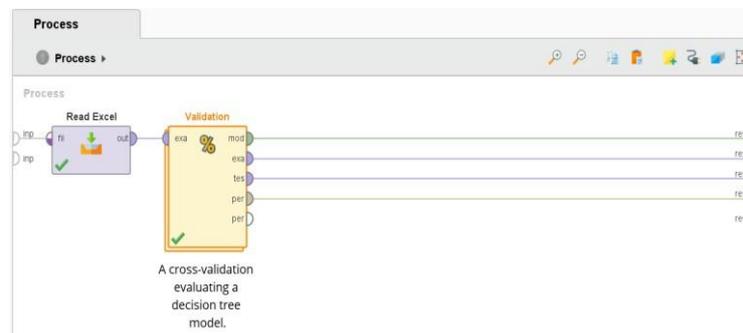
Gambar 6. Hasil Pemodelan

e. Evaluasi

Tahap ini adalah untuk mengevaluasi model prediktif. **10-Fold Cross-Validation** adalah teknik evaluasi model yang digunakan untuk mengukur performa model machine learning secara akurat dan mencegah overfitting. Proses ini melibatkan pembagian dataset menjadi 10 bagian (folds) yang saling eksklusif, di mana setiap bagian digunakan sebagai data uji secara bergiliran, sementara bagian lainnya digunakan sebagai data latih.

Proses 10-Fold Cross-Validation:

- 1) Pembagian Data: Dataset dibagi menjadi 10 subset (fold). Setiap subset memiliki ukuran yang sama, jika memungkinkan.
- 2) Pelatihan dan Pengujian: Proses pelatihan dan pengujian dilakukan sebanyak 10 kali. Pada setiap iterasi:
 - Satu subset (fold) digunakan sebagai data uji.
 - Sisa 9 subset digunakan untuk melatih model.
- 3) Pengulangan: Proses ini diulang hingga setiap subset telah digunakan sebagai data uji sekali. Dengan demikian, setiap data dalam dataset akan diuji.



Gambar 7. Proses Validasi

4) Evaluasi Kegiatan dan Pelaporan

a) Evaluasi

Setelah kegiatan PKM dilaksanakan maka perlu dilakukan evaluasi dengan cara :

PROSIDING SEMINAR NASIONAL

Hasil Penelitian dan Pengabdian Masyarakat

- Mengukur keefektifan dan efisiensi dari pelatihan yang dilakukan.
 - Keberlanjutan Kerjasama dengan mitra akan terus terjalin setelah kegiatan PKM, seperti ikut serta dalam melakukan pelatihan dan pendampingan terkait proses analisis data –data kesehatan.
- b) Laporan
- Tahap terakhir dalam kegiatan ini yaitu menyusun pelaporan mulai dari awal kegiatan sampai tahap evaluasi. Laporan ini dapat digunakan sebagai referensi.

PEMBAHASAN

Penelitian mengenai prediksi pasien diabetes dengan metode data mining telah banyak dilakukan untuk meningkatkan akurasi deteksi dini dan memberikan rekomendasi medis yang lebih tepat. Berikut adalah beberapa aspek utama dalam penelitian ini.

Penelitian dapat menggunakan dataset medis seperti:

- a. Pima Indians Diabetes Dataset (PIDD) dari UCI Machine Learning Repository.
- b. Electronic Health Records (EHR) dari rumah sakit atau klinik.
- c. Data survei kesehatan nasional yang berisi informasi tentang gaya hidup, pola makan, dan riwayat keluarga.

Dataset ini biasanya mengandung fitur seperti:

- a. Usia, indeks massa tubuh (BMI), tekanan darah
- b. Kadar glukosa, insulin, hemoglobin A1c (HbA1c)
- c. Riwayat keluarga, gaya hidup, dan kebiasaan makan

SIMPULAN

Terlaksananya kegiatan PkM ini di dorong oleh beberapa faktor. Pertama, pihak program studi sangat kooperatif. Kedua, adalah dukungan dari para Dosen . Keterbukaan dari Dosen(peserta) untuk menerima pengetahuan baru . Penerapan data mining dalam bidang kesehatan sangat luas dan bermanfaat, terutama dalam menganalisis data medis yang besar untuk meningkatkan kualitas layanan kesehatan, mendukung keputusan klinis, dan mengoptimalkan pengelolaan rumah sakit yaitu untuk melakukan prediksi dan analisis data pasien. Saran untuk kegiatan selanjutnya adalah menjalin kerja sama dengan pihak-pihak lain agar pengetahuan tentang Data Mining lebih luas lagi di terapkan pada bidang kesehatan.

UCAPAN TERIMA KASIH

Terimakasih kepada Yayasan Hang Tuah Pekanbaru melalui Lembaga Penelitian dan Pengabdian kepada masyarakat Hang Tuah Pekanbaru yang memberikan dukungan secara moril dan materil sehingga kegiatan ini dapat terlaksana dengan baik dan lancar.

DAFTAR PUSTAKA

- Binus. (2021, April). *Mengenal Konsep Dasar, Tujuan, dan Teknik Data Mining – School of Information Systems*.
<https://sis.binus.ac.id/2021/04/23/mengenal-konsep-dasar-tujuan-dan-teknik-data-mining/>
- dqlab. (2022). *4 Implementasi Data Science Di Bidang Kesehatan*.
<https://dqlab.id/4-implementasi-data-science-di-bidang-kesehatan>
- Hotz, N. (2024). *What is CRISP DM? - Data Science Process Alliance*. <https://www.datascience-pm.com/crisp-dm-2/>
- Husen, D., Sandi, D., Bumbungan, S., Yogyakarta, U. A., Informatika, M. T., Mining, D., & Forest, R. (2022). *Analisis Prediksi Kebakaran Hutan dengan Menggunakan Algoritma Random Forest Classifier*. 16, 150–155.
- Jiawei, H. (n.d.). *Data mining concepts and techniques - 2006*. Retrieved May 22, 2023,
from <https://elibrary.bsi.ac.id/readbook/221528/data-mining-concepts-and-techniques>
- Kase Sukeza, I. (2022). *CRISP DM Sebagai Salah Satu Standard untuk Menghasilkan Data Driven Decision Making yang Berkualitas*.
<https://www.djkn.kemenkeu.go.id/artikel/baca/15134/CRISP-DM-Sebagai-Salah-Satu-Standard-untuk-Menghasilkan-Data-Driven-Decision-Making-yang-Berkualitas.html>
- Setianingsih, D. (2015). *Penerapan Data Mining dalam Analisis Kejadian Tanah Longsor di Indonesia dengan Menggunakan Association Rule Algoritma Apriori*.
<https://publikasiilmiah.ums.ac.id/handle/11617/5766?show=full>
- Ucha Putri, S., Irawan, E., Rizky, F., Tunas Bangsa, S., -Indonesia Jln Sudirman Blok No, P. A., & Utara, S. (2021). *Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5*. *Kesatria : Jurnal Penerapan Sistem*